# Probabilistic model of cryptographic key collisions in relation to the key length

T. Van Hecke
Ghent University
Faculty of Engineering and Architecture
Voskenslaan 270, 9000 Ghent (Belgium)
Tanja.VanHecke@ugent.be

December 22, 2015

**Abstract**

Cryptographic keys are used to encipher data. Agencies generating these key numbers want to combine a minimal key length for computational and data storage reasons with a guaranteed service of unique key numbers. This paper describes the modeling of the length of the key number to reach an imposed collision improbability.

**Keywords:** cryptography, collision probability, key generation, authenticity

**MSC 2010:** 97P70, 97R70, 97K80, 97K50

## 1 Introduction

Asymmetrical cryptography [4] [5] is recognized to support the current challenges of digital privacy and authenticity [1] over public networks like the Internet. The technology assures the encryption and decryption of data based on a complementary pair of digital keys. Data which was encrypted with one of the keys of the key pair can only be decrypted with the other complementary key. If one of the keys is kept secret (private key) and the other key is made public (public key) to prove the identity of the owner of the private key, privacy and authenticity of data can be assured without the need to transfer private keys over an insecure public network. However in any circumstance the uniqueness of the created key pairs should be assured, avoiding the possibility that two or

more identities may be associated to the same key pair and consequently to the same secured data. The uniqueness depends on the quality of the random key generators and the size of the population of key pairs which is a function of the key length (number of bits used for a key). Within a specific context (e.g. the generation of key pairs for a national electronic identity card) the uniqueness of a key pair can be assured by applying appropriate control mechanisms like verifying if a key has not already been issued in the past (key clash detection). However key uniqueness should also be assured between specific contexts. In practice no control mechanisms exist between such different contexts (e.g. nations mutually verifying the uniqueness of their generated key pairs) and the only way to assure uniqueness is to make it very unlikely that a same key would be generated multiple times. As the number of required key pairs could be considerable (every citizen requiring a set of unique key pairs which are renewed after a certain period of time) the probability of a possible key clash should be known for a certain choice of cryptographic key length. This paper describes a model for estimating the required key length, given a certain probability of encountering a key clash.

## 2 Key length selection

A lot of research has been done on key size selection as a function of hash functions [6]. These are functions that compress input of arbitrary length to a shorter fixed sized output. Sarkar [3] presented a new trade-off between key size and collision probability for universal hash functions. When practical advice needs to be given about the minimum required key size, this is often based on the risk of cracking the encrypted data, extrapolated to the number of years that you need to keep the encrypted data confidential. Lenstra & Verheul [2] use Moore's Law to estimate the required key length to protect data against hostile attacks.

So far, less attention has been spent on the case of collision probability in the random choice of cryptographic keys without duplication check. Therefore, in section 3 we want to model the risk of collision without key clash detection. Section 4 describes the impact on the collision probability when two contexts meet.

## 3 Modeling the requested key length

When $x$ keys have to be selected out of $2^A$ (all possible binary keys of length $A$) by the principle of drawing with replacement, the probability of generating a duplicate key by random generation of keys is called $p_0$ with

$$\begin{aligned} p_0 &= 1 - P(\text{all keys are different}) \\ &= 1 - \frac{\prod_{i=1}^{x-1}(2^A - i)}{(2^A)^{x-1}}. \end{aligned} \tag{1}$$

Given a fixed number $x$ of keys to be selected, the probability (1) can be considered as a function of the length of the key $A$. The shape of the curve depicting $p_0$ as a function of $A$ for $x = 2 \times 10^\alpha$ ($\alpha = 2, 3, 4$) can be seen in Figure 1 and can be modeled by

$$f(A, \alpha) = \frac{1}{2}\,\mathrm{erf}(a + b\,\alpha - \frac{A}{c}) + \frac{1}{2}, \tag{2}$$

with the Gauss error function

$$\mathrm{erf}(x) = \frac{2}{\pi}\int_0^x e^{-t^2}\,dt. \tag{3}$$

A curve fitting procedure gives the best fitting values for $a$, $b$ and $c$, resulting in the model

$$f_0(A, \alpha) = \frac{1}{2}\,\mathrm{erf}(0.6 + 2.7\,\alpha - \frac{A}{2.5}) + \frac{1}{2}. \tag{4}$$

This model for the probability of generating a duplicate key enables us to estimate the requested key length, imposing a collision probability.

## 4 Probability of key collisions between different contexts

Although key collisions can be avoided within one context by preserving the uniqueness of the keys, difficulties can arise when different contexts meet. Two nations for example can use keys of equal lengths $A$ for a national electronic identity card. What is the risk that citizens of different nations use equal private keys for encryption and decryption of confidential data, so privacy can no longer be guaranteed? If $x_1$ keys are well selected in context 1, so no collisions between them are possible and similarly $x_2$ keys are selected in context 2, the chance of intercontextual collisions is

$$p_{ic} = 1 - \left(\frac{2^A - x_1}{2^A}\right)^{x_2}, \tag{5}$$

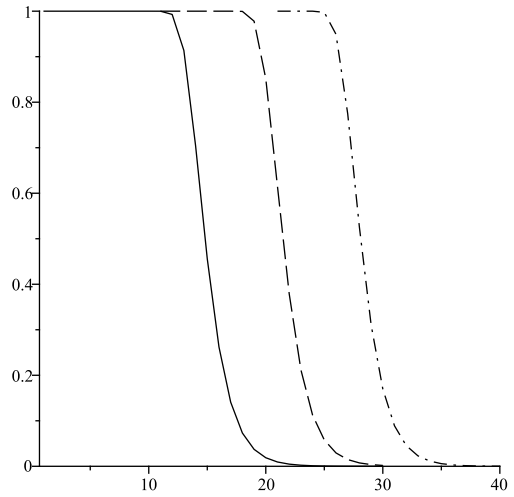


Figure 1: The probability $p_0$ as a function of the length of a key for $x = 2 \times 10^{\alpha}$ ($\alpha = 2$ (solid line), $\alpha = 3$ (dashed line), $\alpha = 4$ (dashdot line)).

Table 1: Collision risk $f_0$ of individual systems and the intercontextual collision risk $p_{ic}$ for a set of key lengths $A$.

| $A$ | $f_0(A,\, 2.8)$ | $f_0(A,\, 3.0)$ | $p_{ic}$ |
|---|---|---|---|
| 30 | $2.8\,10^{-8}$ | $1.5\,10^{-6}$ | $2.3\,10^{-3}$ |
| 35 | $7.3\,10^{-17}$ | $3.3\,10^{-14}$ | $7.3\,10^{-5}$ |
| 40 | $7.2\,10^{-29}$ | $2.8\,10^{-25}$ | $2.2\,10^{-6}$ |

as for all $x_2$ keys $(2^A - x_1)$ good choices can be made out of the possible $2^A$. This is illustrated in Figure 2 which shows that in case of sufficient individual key lengths, a loss of reliability occurs when different contexts meet. The intercontextual collision probability $p_{ic}$ for $x_1 = 2 \times 10^{2.8}$ and $x_2 = 2 \times 10^3$ exceeds the individual collision probabilities $f_0$. Table 1 gives for this example some detailed values by estimating the risk of collisions $f_0$ of both systems and adds the probability $p_{ic}$, estimating the collision risk when those systems meet.

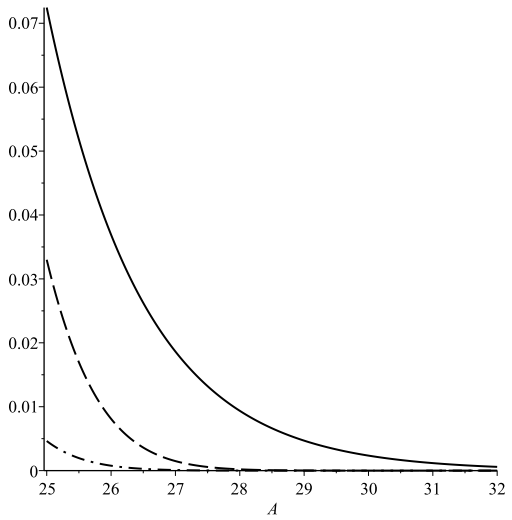


Figure 2: The probability $p_{ic}$ as a function of the length $A$ of a key for $x_1 = 2 \times 10^{2.8}$ and $x_2 = 2 \times 10^3$ (solid line) compared to $f_0(A, 2.8)$ (dashdot line) and $f_0(A, 3)$ (dash line)).

# 5 Conclusions

For random key generation with guarantee of uniqueness, modeling the key length imposes a minimal key length as a function of the required collision probability. Special attention is required when different contexts can meet. Although individual sets of keys can be practically collision free, privacy can no longer be guaranteed when interaction is possible. By modeling this risk, the required enlargement of the key length to approach collision free systems, can be derived.

# References

[1] Banaszak, B. and Rotziewicz K. (2004), Trust and security. Digital citizen cards in Poland, Proceedings of third international conference on electronic government, Zaragoza (Spain), Springer, ed. R. Traunmuller, pp. 342–347.

[2] Lenstra, A. K. and Verheul, E. R. (2001) Selecting Cryptographic Key Sizes, Journal of Cryptology, Vol.14(4), pp.255–293.

[3] Sarkar, P. (2011) A trade-off between collision probability and key size in universal hashing using polynomials, Proceedings of Des. Codes Cryptography, pp.271–278.

[4] Stamp, M. (2011) Information Security: principles and practice, John Wiley & sons.

[5] Trcek, D. (2006) Managing information systems security and privacy, Springer.

[6] https://en.wikipedia.org/wiki/Cryptographic_hash_function